Performance and Robustness Study of Peer-to-Peer Video Streaming Networks

Lyrial Chism, University of Mississippi

Xiaoqing He, University of Minnesota

Liquan Huang, University of Delaware

Ashraf Ibrahim, Texas A&M University

Christopher Jones, University of Pittsburgh

Yan Shu, Georgia Institute of Technology

Mentor: Chai Wah Wu, IBM

August 19, 2008

Abstract

Throughout the last few decades, there have been many advancements in the peer-to-peer based file sharing. Many applications such as BitTorrent and EMule have successfully built a network that distributes files to thousands of users [6]. However, adding video streaming to the peer-to-peer network has been a difficult task. It is believed that the current client-server model of video streaming incurs a significant bandwidth cost for content providers. Therefore, it is not hard to see why using peer-to-peer technology to accomplishing this task would be beneficial. However, peer-to-peer video streaming comes with its own set of challenges. This paper focuses on maximizing the streaming rate of data amongst the source and peers in given networks. We explore this by analyzing different types of networks with a small amount of peers.

1 Introduction

Peer-to-peer file sharing has become a very popular technology for internet users. The server load in peer-to-peer network is much more alleviated than that in the client-server network, which is illustrated in Figure 1.

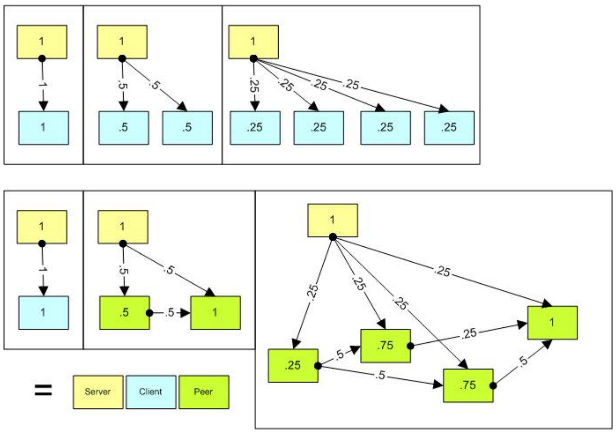

Figure 1: Illustration of a client-server network and a peer-to-peer network [12]

The top structure of Figure 1 is an example of a client-server network. Notice that when the number of clients increases, for a fixed bandwidth for the server (shown in the boxes at the top), the downloading rate between the server and each client decreases. The bottom structure of Figure 1 is an example of a peer-to -peer network. Just as in the client-server network, when the number of peers increases, the downloading rate between server and each peer decreases as well. However, one major difference is that the peers are connected to each other, and thus they can share data amongst themselves. This means that the server is less responsible for distributing data to all the peers and the peers are receiving packets of information from different sources at the same time. Unlike peer-to-peer file sharing, peer-to-peer video streaming has many constraints. One of the most important constraints is real-time streaming [6]. Users expect to watch a downloaded video immediately (with a few seconds of delay). This means that the data must be downloaded quickly in an ordered way. With applications such as BitTorrent, files may take several hours and sometimes days to download and order in which the data is downloaded is not important since the file is not used until it is complete.

The rest of our report is organized as follows. In section 2, we define our problem and formulate it as a linear programming. In section 3, we describe four types of graphs that we are interested in. In section 4, we present our numerical results of the four types of graphs in three scenarios. In section 5, we present some other testing results. In section 6, we explain the implication of the numerical results in section 4 and section 5. In section 7, we discuss future research directions.

2 Maximum Streaming Rate Problem

The maximum achievable rate is defined as the maximum streaming rate at which the source can deliver to all the nodes [3].

2.1 Maximum Achievable Problem

For a given network, we want to find the maximum achievable streaming rate in which the source can download data to all receivers. We are given a graph $G :< N, E >$ of n nodes and an upload bandwidth u_i for each node i. We define s to be the source node (server) and all others to be *peers*. Since users are able to download and upload data at different rates, our graphs are directed. We assume that the download bandwidth of each peer is not a bottleneck and can be assumed to be infinite. $e(i, j)$ is an edge of the graph if data can travel from i to j at some rate r that is no more than u_i. Let f_i be the rate of content that a node i gets from the source such that $f_i \geq r$. Thus, it is the maximum of these r's, r^*, that is the maximum achievable rate. An upper bound for the streaming rate in any bandwidth scenario can be computed as follows: $r^* = \min\left(u_s, \frac{\sum_{i \in N} u_i}{n-1}\right)$ [5]. This rate is achievable for the complete graph.

A packet spanning tree is a directed spanning tree, rooted at the source, that delivers one packet per second [3]. Such a directed tree is also called an s-arborescence. It has been proven that we can attain the maximum achievable rate by finding the total number of packet spanning trees packed in a graph. Hence, $r^* = k^*$ where k^* is the maximum number of packet spanning trees packed into a graph [3].

2.2 The Hardness of Finding Maximum Achievable Rate

Finding maximum achievable rate is a NP-hard problem. For a given connected graph, if the bandwidth of each node is 1, then maximum achievable rate is either 0 or 1. Finding maximum achievable rate in this graph is equivalent to finding a Hamiltonian path starting from the source. Because maximum achievable rate is 1 if and only if there is a Hamiltonian path starting from the source. Since finding a Hamiltonian path in a connected graph is NP-hard problem, finding maximum achievable rate is a NP-hard problem. This argument also shows that if all the bandwidth u_i are equal, then the upper bound in [5] shown above is equal to u_s and it is achievable if there is a Hamiltonian path starting from the source.

2.3 Maximum Streaming Rate

Due the hardness of the maximum achievable rate problem, the authors in [3] give an linear programming relaxation of the maximum achievable rate problem. We use their notation and linear programming formulation as follows. Let x_e be a flow value on edge $e \in E$ and $f_{i,e}$ be the amount of flow node i gets from the source via edge e.

$$\begin{aligned}
\text{maxmize} \quad & \mathit{StreamRate} \\
\sum_{e(s,j)\in E} f_{i,e} - \sum_{e(j,s)\in E} f_{i,e} &= \mathit{StreamRate} \quad \forall i \in N \setminus \{s\} \\
\sum_{e(j,i)\in E} f_{i,e} - \sum_{e(i,a)\in E} f_{i,e} &= \mathit{StreamRate} \quad \forall i \in N \setminus \{s\} \\
\sum_{e(b,a)\in E} f_{i,e} - \sum_{e(a,c)\in E} f_{i,e} &= \mathit{StreamRate} \quad \forall i \in N \setminus \{s\}, \forall a \in N \setminus \{i, s\} \\
x_e - f_{i,e} &\geq 0 \quad \forall e \in E, i \in N \setminus \{s\} \\
\sum_{e(i,j)} x_e &\leq u_i \quad \forall e \in N \\
f_{i,e} &\geq 0 \quad \forall e \in E, i \in N \setminus \{s\}
\end{aligned}$$

In our study, $f_{i,e} = 0, \quad \forall e(j, s) \in E$ since we assume that the source does not download from the peer. Notice that the optimal value of the above linear programming which we refer to as maximum streaming rate is a good approximation of the maximum achievable rate for their difference is at most 2[1]. This is a starting point of our work in this report. Later on, we implement this linear programming in Matlab and use GNU linear programming kit (GLPK) to solve the optimal value for specific graphical networks. In the next section, we introduce four types of graphs that we conduct tests on.

3 Types of Networks

3.1 Random Graphs

Random graphs are graphs formed by a stochastic process. $G(n, p)$ is a random graph with n nodes that the probability of occurrence of an edge between any two nodes is p. Random graph theory comprises the earliest study of complex networks with theoretical foundations dating back to the 1950s. Only connected random graphs are of interest in this paper (as the maximum streaming rate is 0 otherwise).

3.2 Small World Graphs

Small world graphs are used to model social networks. Though all of the nodes are not adjacent to one another such as in a complete graph, each can be connected to another by a reasonably short path. These graphs are generated in a number of ways. This paper focuses on two models.

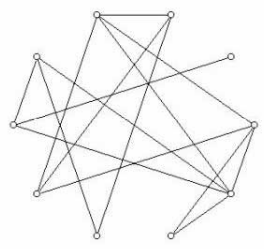

Figure 2: Example of Random Graph [9].

3.2.1 Watts and Strogatz Model

In Watts and Strogatz model, edges are added at random to a cycle of n nodes and never deleted. See Figure 3 for Watts and Strogatz small world graph.

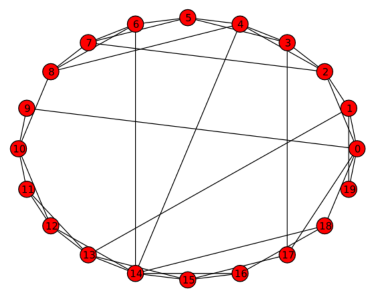

Figure 3: Example of a Watts and Strogatz small world graph [8].

3.2.2 Newman-Watts Small World Graphs

The Newman-Watts small world graphs begin similarly to the Watts and Strogatz Model. However, this one differs in that edges of the original cycle graph are deleted while random edges are being added. We consider two special cases for this model. In case 1, we begin with a simple circle, and in case 2, nodes are still arranged in a circle but all nodes in the starting graph are connected to four nearest neighbors rather than two in the cycle graph. The following figure shows the dynamic process when we rewrite edges randomly in case 2. Notice the differences as the graph changes in Figure 4.

3.3 Scale-free Graphs

The scale-free model starts with a small network in which all nodes are connected to each other. For every node that is added, a fixed number of edges

Figure 4: Example of a Newman-Watts small world graph in case 2 [9].

is added to connect this node to some other nodes already in the graph. The selection of nodes to which the new node is connected is based on the node degree proportion. Scale-free graphs attempt to capture features of large scale graphs both manmade and naturally occurring. Figure 5 is a small example.

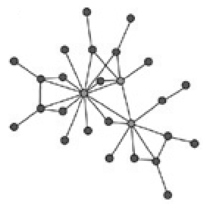

Figure 5: Example of a scale-free graph [11].

3.4 Grid Graphs

A grid graph is an m dimensional $n \times n$ graph such that G is the graph Cartesian product $P_n \times P_n \times ... \times P_n$ (m times)[7]. Figure 6 is a 2-dimensional grid graph.

4 Numerical Results and Corresponding Analysis

Using the four types of graphs in previous section as models for peer-to-peer network, we consider the following three bandwidth scenarios:

1. All peers and source have the same upload bandwidth($u_i = 10$), we shall refer to this scenario as the *homogeneous* scenario.

2. All peers have a constant upload bandwidth ($u_i = 10$) and the source has an upload bandwidth 100 times as large ($u_s = 1000$), we refer to this scenario as the *strong source* scenario.

3. The peers and source have an upload bandwidth taken as samples from a realistic bandwidth distribution. The following data about the user

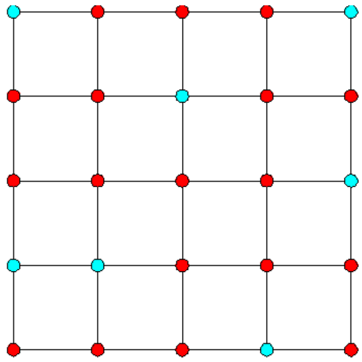

Figure 6: Example of a 5×5 grid graph [10].

upload bandwidth distribution in a real life video streaming application was found in [4].

	modem	ISDN	DSL 1	DSL 2	Cable	Ethernet
Upload	64	256	128	384	768/384	768
share (%)	2.8	4.3	14.3	23.3	18.0	37.3

For the “Cable”upload bandwidth, we used the average of the two upload values given, 576 in our test. We refer to this scenario as the *random bandwidth* scenario.

4.1 Random Graphs

4.1.1 Random Graphs with Homogenous Bandwidths

A result of Erdős and Rényi [2] states that the threshold for random graph of n nodes to be connected is given by

$$p = \frac{\ln(n)}{n}.$$

Fixing n and varying p, we obtain a relative completeness curve for random graphs with n nodes. Clearly, as p approaches 1, any particular $G(n, p)$ is more likely to be a complete graph. Likewise, by the Erdős-Rényi threshold, as p approaches 0, any particular $G(n, p)$ will less likely be complete, indeed for $p < \frac{\ln(n)}{n}$, the graph will like to be disconnected.

Along this research line, for fixing node number n and varying edge probability p, we study the ratio k/c, where k is the maximum streaming rate of random graphs ,and c is the maximum streaming rate of complete graph. Since our graph is randomly generated, we fixed node 1 to be the source. We can find that the convergence of k/c to 1 is faster for larger n as depicted by the graph in Figure 7.

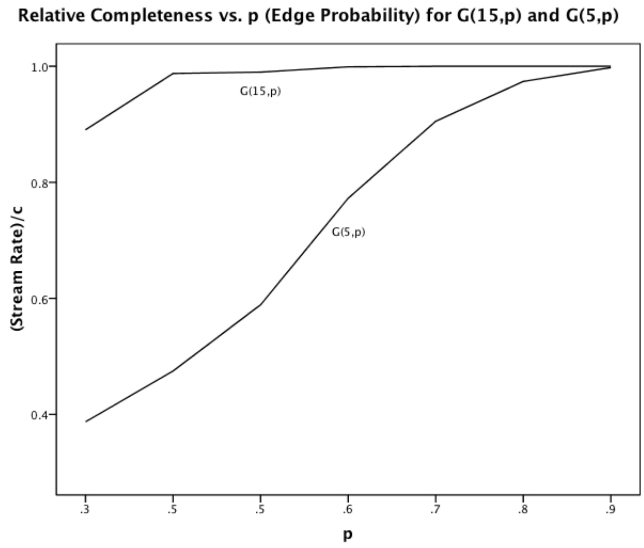

Figure 7: Relative completeness with respect to edge probability p

4.1.2 Random Graphs with Strong Source Bandwidths

Figure 8 shows the maximum streaming rate distribution for $p = 0.5$ and Figure 9 shows the maximum streaming rate distribution for $p = 0.3$. Notice that in both cases there is a skewness which suggests that we are more likely to obtain a streaming rate higher rather than below the mean. Next, we look at the case $p = 0.8$ pictured in Figure 10. Considering what we have observed with the relative completeness curve, we expect these graphs to be close to being complete and thus have a bias towards the maximum streaming rate.

4.1.3 Random Graphs with Random Bandwidths

For random $G = G(n, p)$, we define randomly deleting $q \cdot 100$ percent of the edges in G as a decay iteration of G. For a random graph with random bandwidths, we plot the changing of its maximum streaming rates in 11 decay iterations.(Figure 11)

4.2 Small World Graphs

4.2.1 World Graphs with Homogenous Bandwidths

In the Watts and Strogatz Model, the result is always the maximum achievable rate which was 10. The reason is that the small world graph of this model always contains a Hamiltonian path from the source, along which the network can achieve its maximum streaming rate.

In Newman-Watts Small model, we consider two cases. Case 1. In this model, we begin with a circle, i.e., nodes are only connected to two neighbors. We fix the node size in these trials as 30 and investigate the probability of

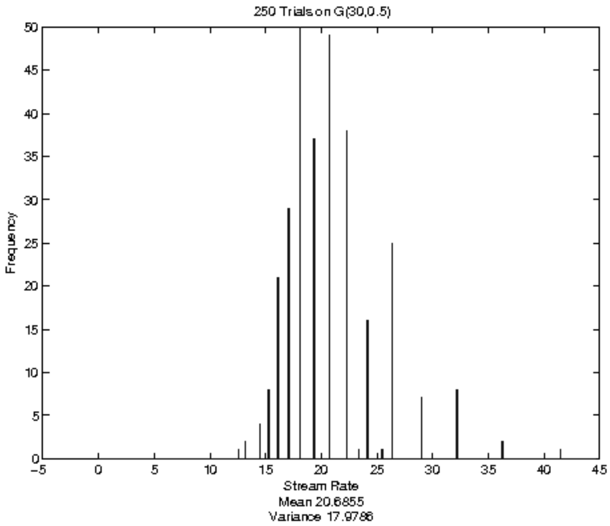

Figure 8: Distribution of streaming rate for $p = 0.5$

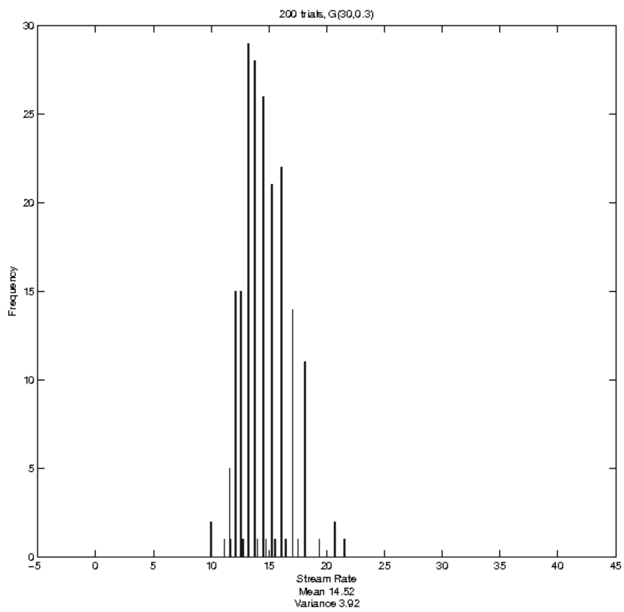

Figure 9: Distribution of maximum streaming rate of random graph $G(30, 0.3)$

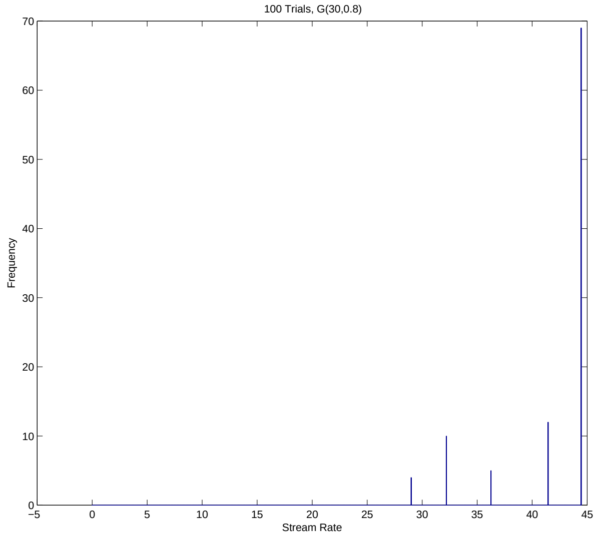

Figure 10: Distribution of maximum streaming rate of random graph $G(30, 0.8)$

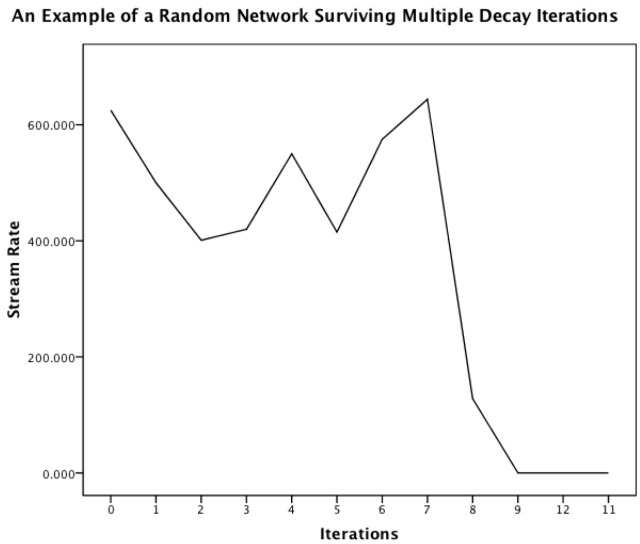

Figure 11: Random graph with random bandwidths in 11 decay iterations

connectivity after adding and removing edges randomly. The result is illustrated in Figure 12.

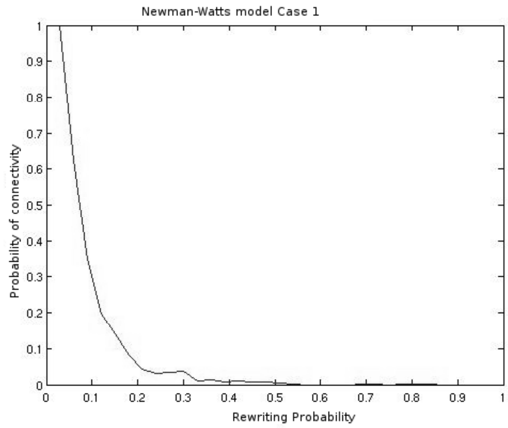

Figure 12: Probability of connectivity w.r.t. rewriting probability in case 1

Case 2. Now, we will show the probability of connectivity when nodes are originally connected to four neighbors. In the illustration below, notice in Figure 13 that even when we rewrite the edges with a probability of 90%, the connectivity probability does not decrease much.

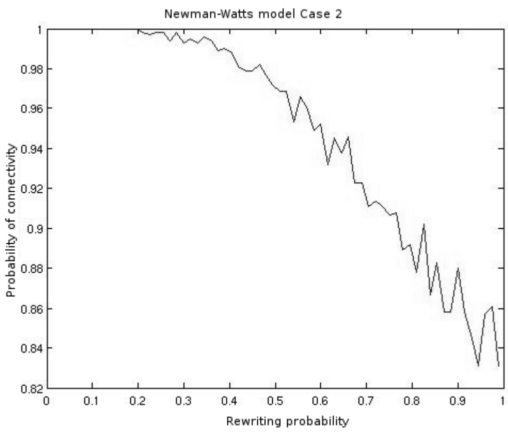

Figure 13: Probability of connectivity w.r.t rewriting probability in case 2

We then test the maximum streaming rates of the resulting graphs with different probabilities in both cases. For case 1, from Figure 14, it is easy to see that even though the rewriting probability is only $p = 0.10$, half of the streaming rate is 0. That's because most of the resulting graphs after rewriting are not even connected.

In contrast to case 1, most of the maximum streaming rate of the resulting graphs is 10 in case 2 which is illustrated in Figure 15.

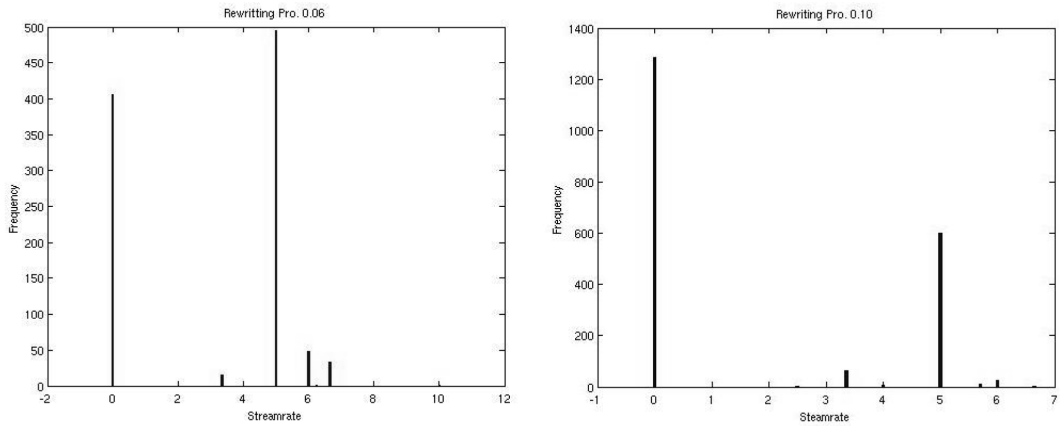

Figure 14: Distribution of maximum streaming rates in case 1

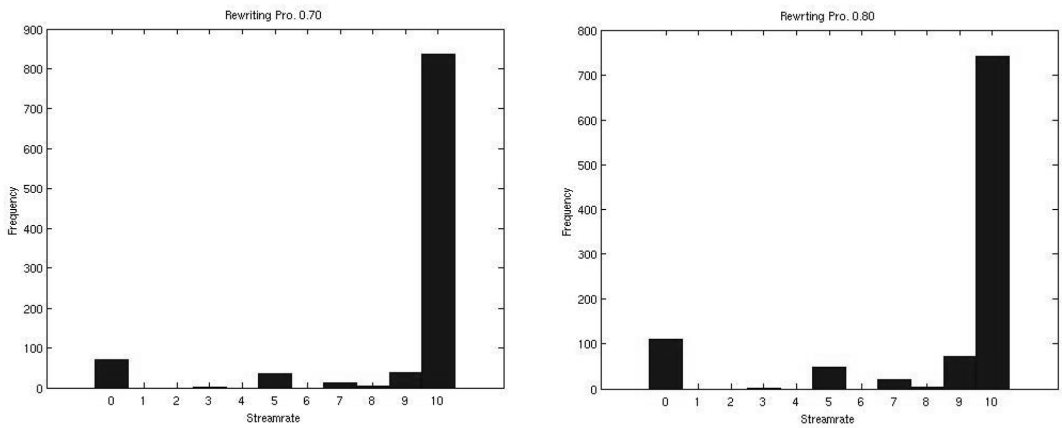

Figure 15: Distribution of maximum streaming rates in case 2

4.2.2 Newman-Watts Small World Graphs with Strong Source Bandwidths

We apply different probabilities in case 1 and case 2 of the Newman-Watts small world graphs. We get the similar results as in the homogeneous bandwidths. That is, in case 1, many maximum streaming rates are still 0; in case 2, most of the maximum streaming rates are around 10.(see Figure 16 and Figure 17).

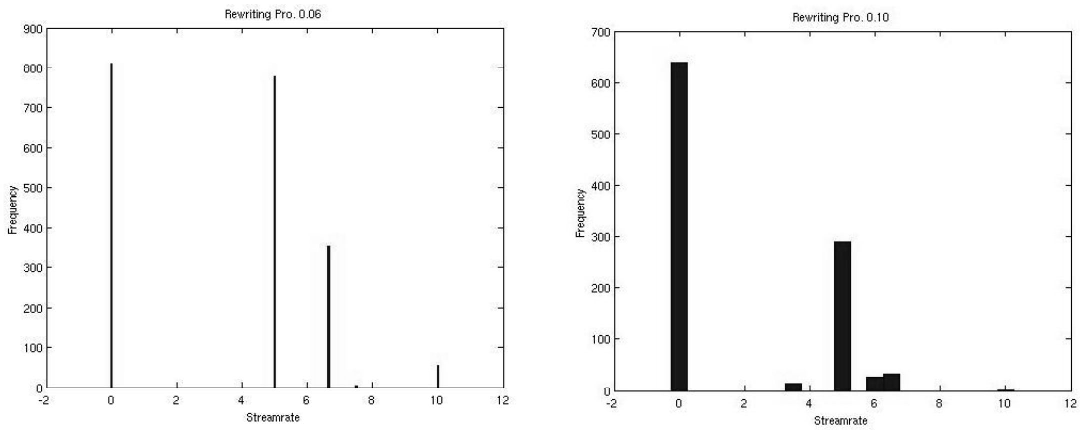

Figure 16: Distribution of maximum streaming rate in case 1

4.2.3 Newman-Watts Small World Graphs with Random Bandwidths

It is useful to scale down the bandwidth vector to have a narrow range of streaming rates. In our test, we scale down the bandwidth to $\frac{1}{7}$ of its original values. In case 1,we still test on two rewriting probabilities, namely $p = 0.06$ and $p = 0.10$. See Figure 18.

We also examine case 2 of Newman-Watts model in which there are four neighbors for each node. Look at the differences between the Figure 18 and Figure 19.

4.3 Scale-free Graphs

The scale-free graphs used for our trials in the following subsections are all generated in the same way. We start with a complete graph of 5 nodes. In each of the following steps, we add a new node and attach it to three nodes in the previous graph. The selection of the three nodes is based on the node degree proportion. We stop when the graph has exactly 30 nodes.

4.3.1 Scale-free Graphs with Homogeneous Bandwidths

For homogenous bandwidths $u_i = 10$, we generate 300 scale-free graphs and calculate their maximum streaming rates. Figure 20 is a plot of the distribution of these maximum streaming rates. It shows us that a large portion of the distribution of the streaming rates is 10.

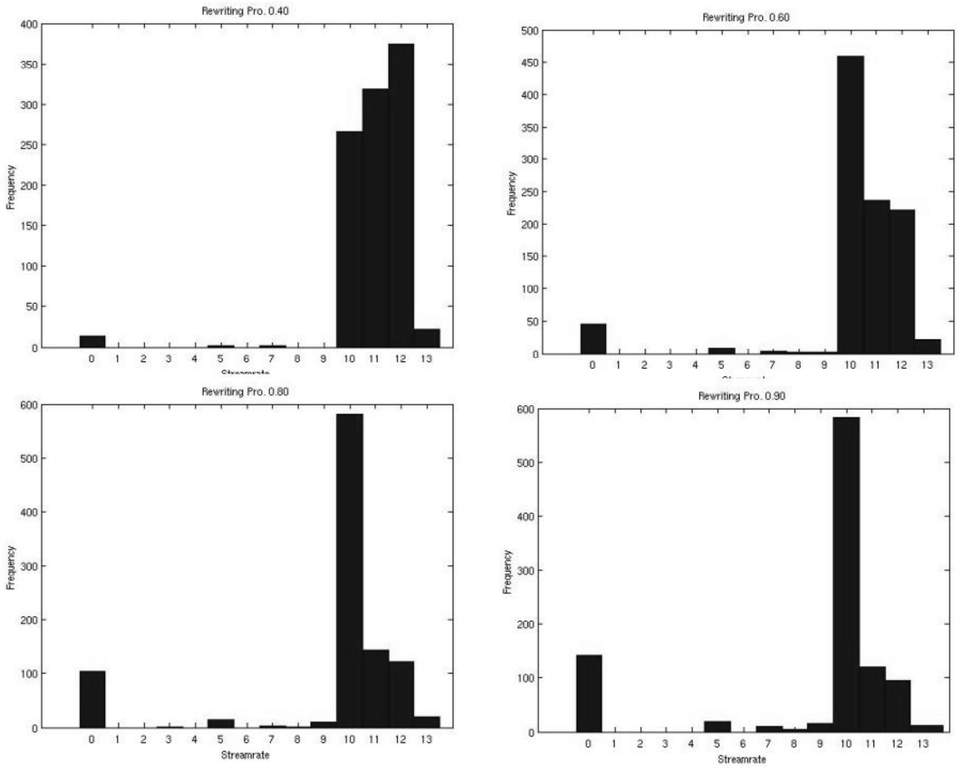

Figure 17: Distribution of maximum streaming rate in case 2

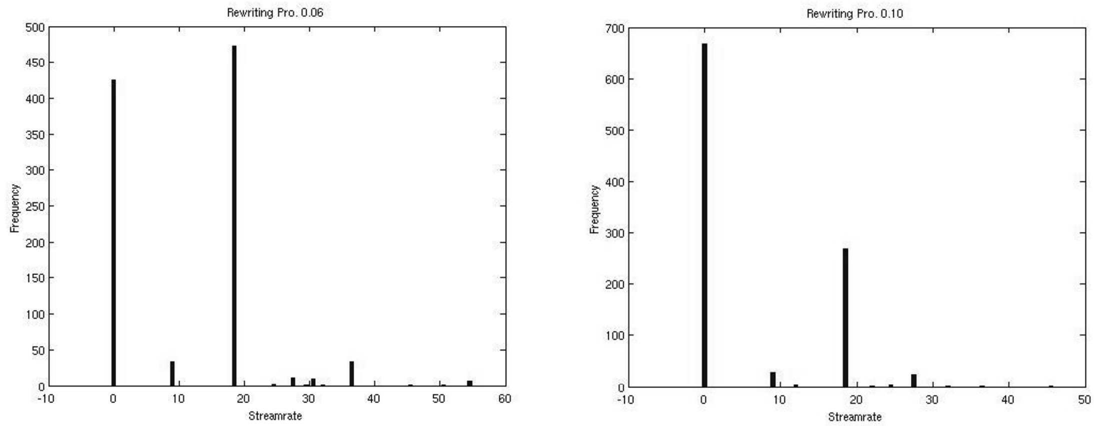

Figure 18: Distribution of maximum stream rate in case 1

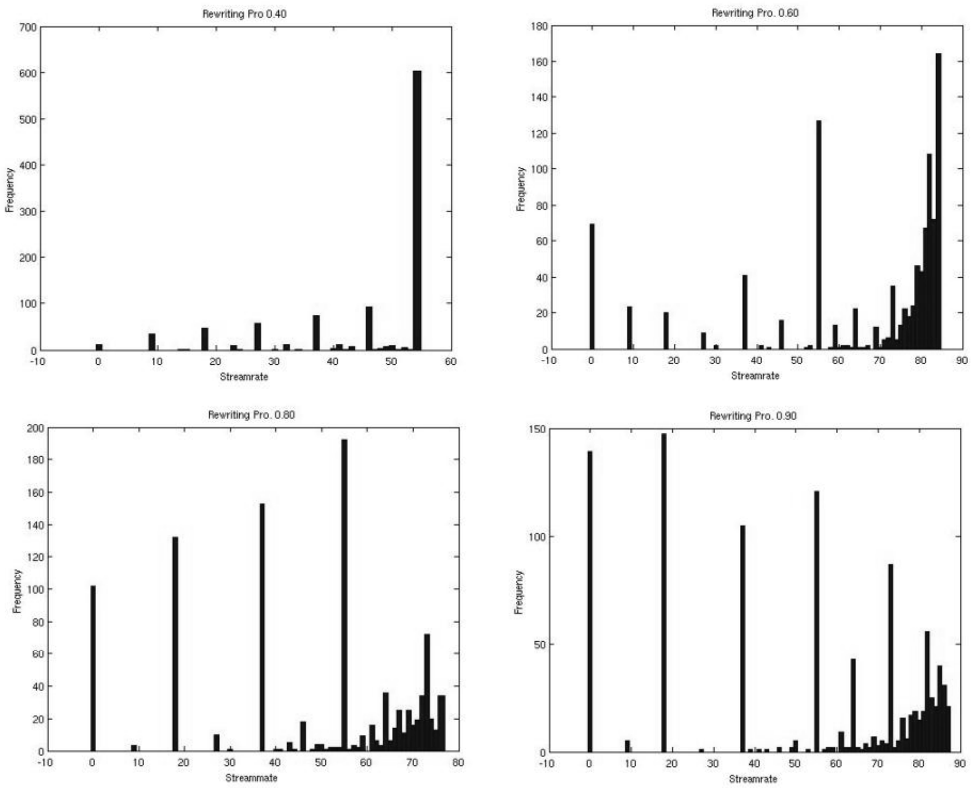

Figure 19: Distribution of maximum stream rate in case 2

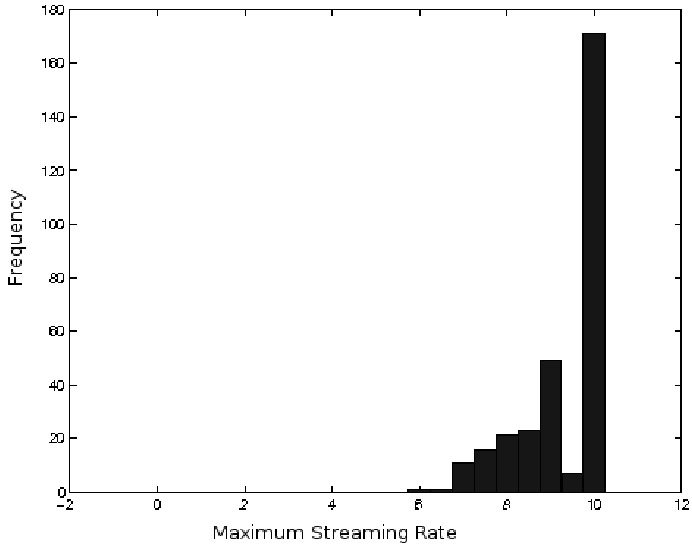

Figure 20: Distribution of Maximum Streaming Rates

4.3.2 Scale-free Graphs with Strong Source Bandwidths

For strong source bandwidths, we also generate 300 scale-free graphs. The mean of the 300 maximum streaming rates is 10.7318 and the maximum of maximum streaming rates amongst this trial is near 19. Please refer to Figure 21 for the more detailed result.

4.3.3 Scale-free Graphs with Random Bandwidths

For the random bandwidths case, we still generate 300 scale-free graphs. Their maximum streaming rates range from 300 to 600(see Figure 22).

4.4 Grid Graphs

In this subsection, we focus on 5×5 grid graph. For each node, we calculate the maximum streaming rate of taking this node as source and all the other nodes as peers.

4.4.1 Grid Graphs with Homogenous Bandwidths

In the homogeneous bandwidth case, we find that the maximum stream rates alternate between 9 and 10 when the source node change from one position to its neighborhood, which is shown in the Figure 23. The interesting phenomenon comes from the fact that for any two adjacent nodes in the grid graph, exactly one of them has a Hamiltonian path starting from it.

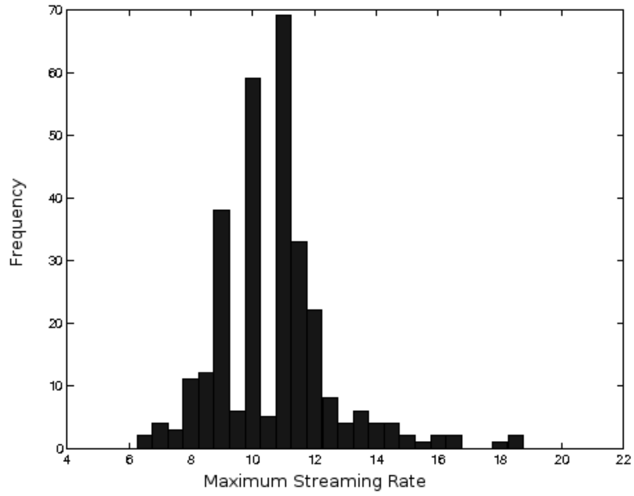

Figure 21: Distribution of maximum streaming rate

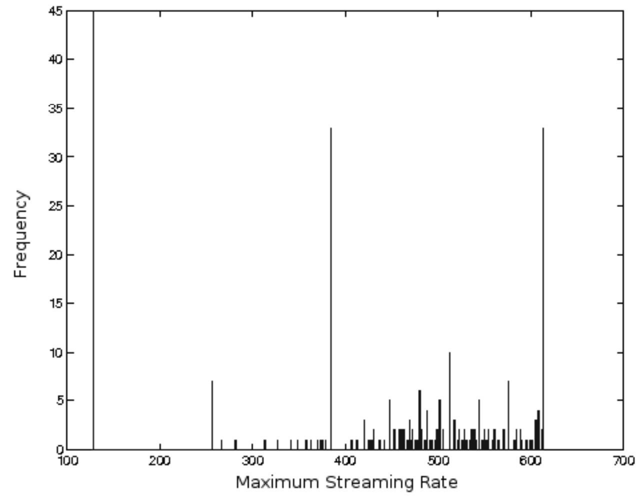

Figure 22: Distribution of maximum streaming rate

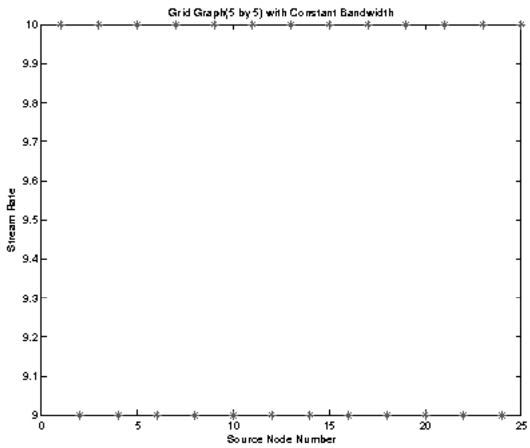

Figure 23: Distribution of maximum streaming rate of 5×5 grid graph with homogeneous bandwidth

4.4.2 Grid Graphs with Strong Source Bandwidths

We set the bandwidth of the source node as 1000 and the bandwidth of all the other nodes as 10. In the first trial, we take the first node on the boundary as the source. In the second trial, we choose the center node as the source. Surprisingly, in both situation, we obtained the same result as in the homogenous bandwidths case.

4.4.3 Grid Graphs with Random Bandwidths

For randomly generated bandwidths, we generate two samples each of which consists of 100 tests. For each node, we average of the maximum streaming rates of taking this node as source. Just as in the constant bandwidth case, despite which node is the source, the average maximum streaming rate does not change much. In Figure 24, you can see that the average maximum streaming rates vary around 350.

5 Other Testing Results

5.1 Random Graphs

We look at robustness of random graphs in the strong source scenario. All figures below represent 20 percent decay iterations. Figure 25 shows how the maximum streaming rates changes after successive decay iterations.

5.2 Watts and Strogatz Model Small World Graphs

For Watts and Strogatz Model small world graphs, we also track the maximum streaming rate as edges are added. Thus, we start with 10 nodes and randomly

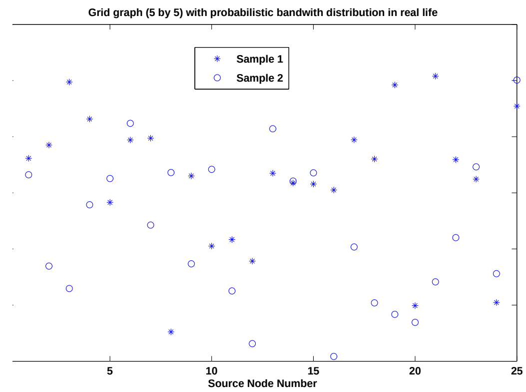

Figure 24: Distribution of average maximum streaming rate of 5×5 grid graph with random bandwidths

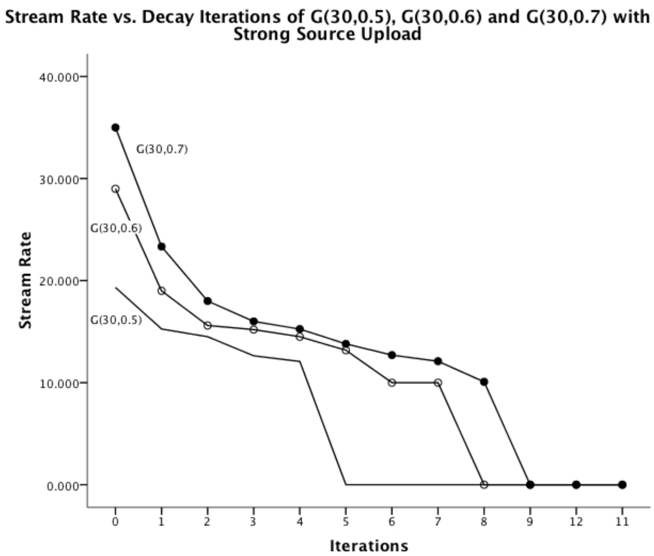

Figure 25: Robustness of random graph with strong source bandwidth

add edges. The bandwidths are randomly chosen from integers between 1 and 20. Figure 26 illustrates the relationship between the performance and the total number of edges.

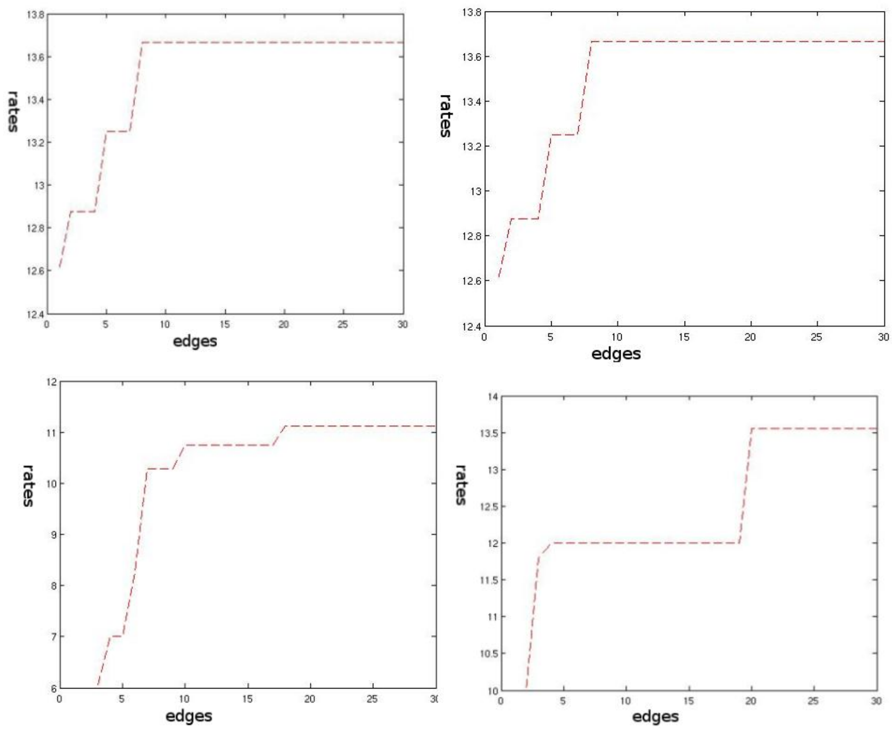

Figure 26: Maximum streaming rates with random bandwidths

5.3 Scale-free Graphs

For this model, we also study how the maximum streaming rates change as the number of nodes increase for a scale-free graph. We start with 5 nodes, add 3 edges per new node, and add a total of 25 nodes. Each time a node is added, we calculate the maximum streaming rate. We average the maximum streaming rates for 300 scale-free graphs. The streaming rates are illustrated in Figure 27.

Due to limitations to the number of trials, the curve is not smooth. However, it decreases as the number of nodes increases. We do this test again for 200 graphs and change the bandwidth of the source node to be 1000. Figure 28 shows our result.

Notice that the rate decreases very quickly at the beginning. When the number of nodes increases to 6, the rate drops from about 160 to about 20. This is because that we start from a complete graph which has relatively higher rates than other graphs with the same number of nodes. Later, the rate drops much slower. In addition, we also consider the random bandwidth scenario. We only track the addition of 15 nodes per graph for 200 trials due to the limitation

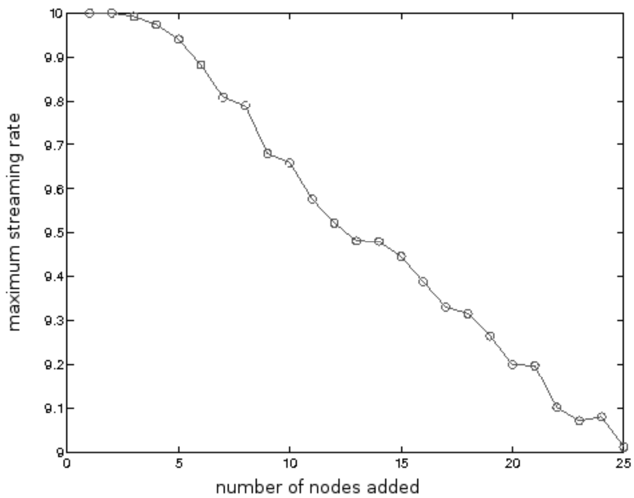

Figure 27: Changes of maximum streaming rates as number of nodes increases k

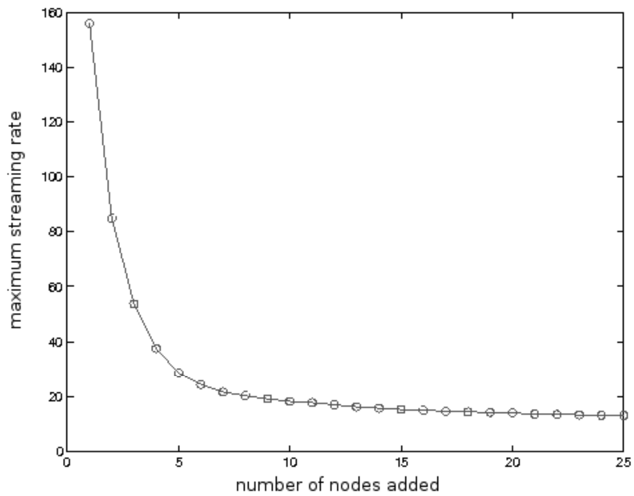

Figure 28: Changes of maximum streaming rates as number of nodes increases

of time(Figure 29). After adding the 15 nodes, the mean of the maximum streaming rates drops by approximately 200.

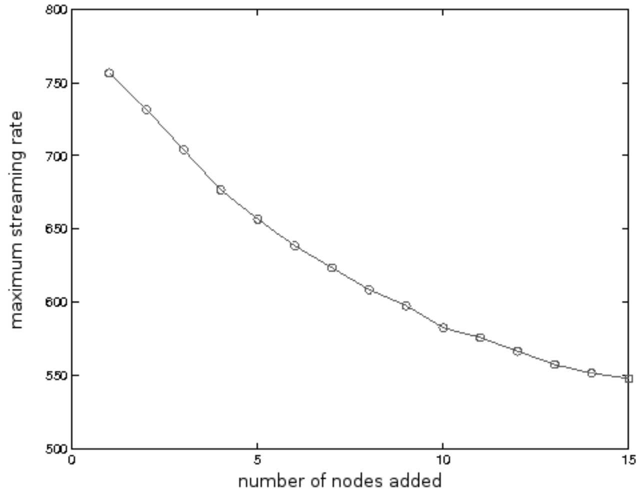

Figure 29: Changes of maximum streaming rates as number of nodes increases

5.4 Grid Graphs

Because grid graphs can be defined in various dimensions, we did some tests on grid graphs in 3 and 4 dimensions as well. Figure 30 is a comparison of the maximum streaming rates in three different dimensions.

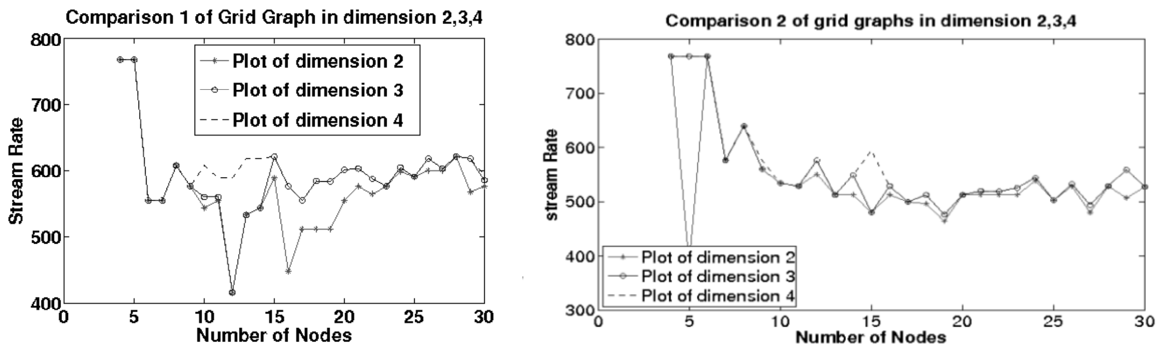

Figure 30: Comparison of grid graph performance in dimension 2,3,4

Then we study 5×5 grid graph with different bandwidth distribution. We divide the grid graphs into three areas: center node, interior nodes excluding the center, and boundary nodes. The first test is to assign nodes with bandwidth decreasing from the center to the boundary. That's, the bandwidth for the center was 11, interior nodes, 7, and boundary nodes 3. In the second test, we assign nodes with bandwidths increasing from the center to the boundary. The bandwidth for the center was 3, interior nodes, 7, and boundary nodes 11. For each node, we calculate the maximum streaming rate of taking this node as source and all the other nodes as peers. When the bandwidth is decreasing from the center to the boundary, the streaming rates range from 3 to 4. In the increasing case, the streaming rates range from 7 to 9 with the exception of the center. The results are also plotted in Figure 31.

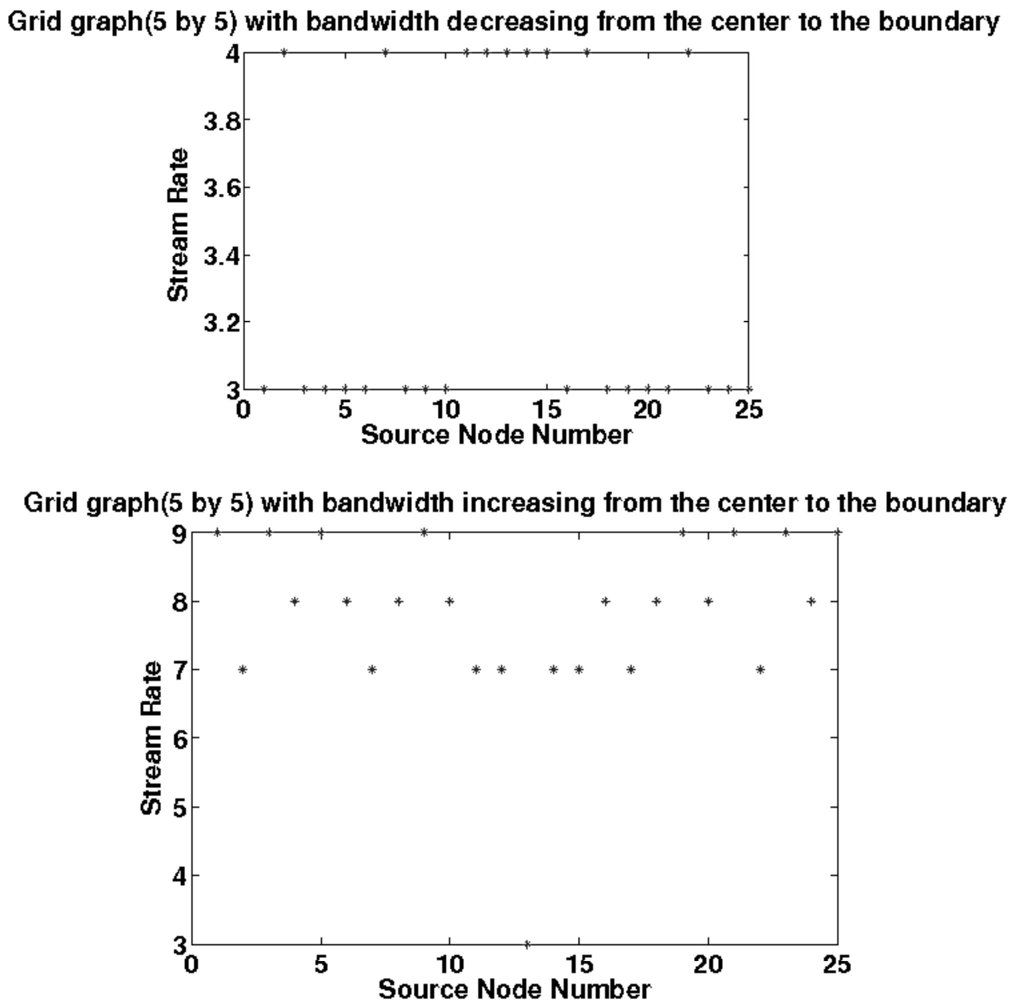

Figure 31: Comparison of maximum streaming rate with different bandwidth distribution

6 Conclusion

6.1 Random Graphs

For random graph $G(n, p)$, the ratio of the maximum streaming rate of G over that of the complete graph of n nodes converge more faster to 1 as p approaches 1 if we increase n.

6.2 Small World Graphs

6.2.1 Watts and Strogatz Model

In the case of the more dynamic small world topology, better streaming rate performances are attained with higher means of bandwidth distribution. The highest possible rate, which is usually no more than 30% of the source's bandwidth, that can be achieved seems to be accomplished only when the bandwidth of the peers is less than 30% of the source's bandwidth. This does not seem to be the most efficient topology amongst the tested subjects.

6.2.2 Newman-Watts Model

We have noticed that as long as the graphs are connected after rewriting edges, then the maximum streaming rates are highly concentrated. These streaming rates are also within relatively narrow ranges.

6.3 Scale-Free Graphs

When the bandwidths are constant or do not differ by tremendous amounts, the scale-free graph is a very efficient model for peer-to-peer networks. We have concluded that with a smaller range, we get a better maximum streaming rate. When the bandwidth of the source is much greater than the bandwidth of the peers, the maximum streaming rate is constrained. As for our dynamic trials, the maximum streaming rate always decreases. In particular, it decreases very quickly for the beginning nodes. Thus, scale free are ideal in certain networks.

6.4 Grid Graphs

The grid graph does not seem to be the ideal model for a peer-to-peer network. Even when the bandwidth at the source is much greater than the bandwidth of the peers, the maximum streaming rate is not large. One noticeable feature of these graphs is that they are stable in the sense that the choice of the source node does not affect the streaming rate. We must also note that the maximum streaming rate converges in any dimension. This means that for a large sample of nodes, the maximum streaming rate will not change greatly from if we change the graph topology from dimension to dimension.

7 Future Work

Some questions to answer in future work are the following: Given the dependence of the streaming rate and its robustness on the network topology, how should this guide the way new peers are joined into a peer-to-peer network? For local bandwidth allocation algorithms which are suboptimal, do they exhibit the same dependence on the network topology?

Acknowledgements

This report could not be finished without the help of people who we would like to express our gratitude here. First, we would like to thank IMA for providing us with this opportunity and such a great working environment. In particular, we thank the workshop organizers, Professors Fadil Santosa and Richard Braun and their staff for making the workshop as stimulating and enjoyable as possible. We thank IBM for giving us the chance to do this project. We also would like to extend our thanks to the other teams, especially their mentors.

References

[1] N.Bansal, R.Khandekar, and V.Nagarajan, "Additive Guarantees for Degree Bounded Directed Network Design", In *STOC*, 2008

[2] B. Bollobas. *Random Graphs.* Sept 2001.

[3] Z. Fu. "Throughput Bound in P2P Streaming System in General Overlay Topology." Submitted for publication.

[4] C. Huang, J. Li, and K. Ross. "Can Internet Video-on-Demand be Profitable?" In *ACM SIGCOMM*, Aug 2007.

[5] R. Kumar, Y. Liu, and K. Ross. "Stochastic Fluid Theory for P2P Streaming Systems." In *IEEE INFOCOMM*, 2007.

[6] B. Li, J. Liu, S. Rao, and H Zhang. "Opportunities and Challenges of Peer-to-Peer Internet Video Broadcast." In *Proc of IEEE*, vol 96, pp 11-24, 2008.

[7] E. Weisstein. "Grid Graph." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/GridGraph.html

[8] http://en.wikipedia.org/wiki/Watts_and_Strogatz_model

[9] http://epress.anu.edu.au/cs/html/ch05s03.html

[10] http://www.mts.jhu.edu/~ers/matgraph/samples/html/domination.html

[11] http://www.nature.com/nrg/journal/v5/n2/box/nrg1272_BX2.html

[12] http://www.visiognomy.com/diagrams/archives/2005/02/03/clientserver-vs-p2p/